A MACHINE LEARNING-BASED LIFE CYCLE ASSESSMENT PREDICTION MODEL FOR THE ENVIRONMENTAL IMPACTS OF BUILDINGS

Ann Francis[1] and Albert Thomas

Civil Engineering Department, IIT Bombay, Powai, 400076, India

With the emerging importance of achieving climate targets and net-zero levels, assessing the environmental sustainability of buildings is of paramount importance. Life Cycle Assessment (LCA) is a popular tool used for such assessment. However, performing LCA for buildings is time-consuming and challenging due to inconsistencies in the databases, software limitations, and data intensiveness, making it a complex tool for decision-making applications. Therefore, this study proposes a methodological framework to develop surrogate LCA models for buildings using modern machine learning (ML) tools such as Multiple Regression and Artificial Neural Networks (ANN). Such a framework improves the application of LCA in environmental decision-making during the planning of building projects by reducing the time, effort, and complexity associated with conducting LCA of buildings. It can be found that the mean absolute percentage error (MAPE) for the tested dataset in the regression-based model is less than 5 percent rendering it a good surrogate model.

Keywords: life cycle assessment; artificial neural networks; multiple regression

INTRODUCTION

Buildings significantly influence the environment due to large resource consumption, energy and water use, and emission and waste generation associated with construction, operation, and end of life (Francis and Thomas, 2020). Therefore, it is essential to balance out these impacts through improved decision-making and policies to achieve global targets of net-zero and sustainable development goals. Although several tools were developed for environmental assessment, Life Cycle Assessment (LCA) is the most popular and standardized tool in this domain. LCA helps quantify the effects of human activities on the environment (Barros and Ruschel, 2020). However, buildings have a very long lifespan and are subject to changes in building characteristics, particularly during their operational phase (Fouquet *et al.,* 2015). Hence, modern research is transforming its outlook from a static to a dynamic environmental assessment of buildings (Beloin-Saint-Pierre *et al.,* 2020; Francis and Thomas, 2022a; Levasseur *et al.,* 2010).

However, the dynamic and diverse behaviour of buildings and the data intensiveness associated with their scale and size creates a lot of complexity and variability in environmental assessment (D'Amico *et al.,* 2019). Further, the absence of standard

[1] annfrancis@iitb.ac.in

Francis, A and Thomas, A (2022) A Machine Learning-Based Life Cycle Assessment Prediction Model for the Environmental Impacts of Buildings *In:* Tutesigensi, A and Neilson, C J (Eds) *Proceedings of the 38th Annual ARCOM Conference,* 5-7 September 2022, Glasgow, UK, Association of Researchers in Construction Management, 592-601.

environmental databases and non-uniformity in assessment methods makes buildings a more challenging product for evaluation as it comprises of several materials, transportation, and significant energy and water consumption (Morales *et al.,* 2020). Further, decision-making and policy analysis in the building sector involves evaluating numerous alternatives and scenarios to arrive at the best possible solutions in terms of environmental sustainability. However, the rigorous data inventory requirement, data gaps due to improper reporting of possible impacts, and the non-uniformity of methods and databases make LCA a complex tool for such elaborate decision-making applications in the building sector.

To address the challenges mentioned above associated with conducting LCA of buildings and utilising its potential in decision making, it is worth exploring the potential of modern ML tools in this domain. Such modern computational tools of simulation and ML could ease the complexities of applying environmental assessment tools. Studies have demonstrated that such tools can be utilized to develop surrogate prediction models that enable efficient, robust, and faster analysis (Sousa *et al.,* 1999; Ziyadi and Al-Qadi, 2019).

Therefore, this paper proposes a framework for developing an ML-based LCA prediction model to perform LCA for buildings. It demonstrates the application of ML tools such as multiple regression algorithms and Artificial Neural Networks (ANN) to create robust surrogate models to predict the environmental performance of buildings. Such prediction models save time and reduce the complexity of environmental analysis while enabling efficient and dynamic assessment of the built environment. Such models enable LCA application for better environmental decision-making of buildings by enhancing its flexibility and capability to evaluate numerous scenarios and alternatives while planning building projects.

LCA helps quantify the environmental impacts of a product or a system (Rebitzer *et al.,* 2004). It takes a cradle to grave approach by considering the life cycle effects of a product or activity from the procurement of raw materials to its final demolition and disposal. International Organisation of Standardisation (ISO) standards are available to perform LCA (Guinee *et al.,* 2011). The existing studies show that LCA has been popularly applied in the built environment for research and practical applications (Cabeza *et al.,* 2014; Chau *et al.,* 2015; Sharma *et al.,* 2011). However, the widespread application and use do not negate that LCA is a time-consuming method with complex computing and analysis involved (Sousa *et al.,* 1999). Notably, when buildings are the products analysed, this complexity increases with the long-life span and varying dynamic nature of building characteristics, such as changing properties of building materials, surroundings, and energy consumption. Hence, as generally done in a conventional LCA, a static assessment would underestimate the actual impacts due to buildings (Francis and Thomas 2022a). Data gaps and non-uniformity of databases are challenges in performing dynamic LCA for buildings (Hellweg *et al.,* 2014). These challenges limit the application of LCA-based decision-making during the planning of building projects.

However, with the growing necessity to achieve climate targets, sustainable development goals, and eventually net-zero targets, it becomes necessary that environmental performance is given primary importance while planning and decision making of building projects (Francis and Thomas, 2022b). This demands the application of modern ML tools to address these complexities. ML tools are used to develop substitutable prediction models which output the results that closely represent

the theoretical values. Such surrogate models are then used for prediction purposes due to their application simplicity and flexibility enabled for constantly changing evaluation parameters. Developing ML-based prediction models requires collecting extensive input/initial data. This data, after pre-processing, is then trained using a machine-learning algorithm. A suitable supervised/unsupervised learning algorithm will study the pattern of this input data and develop a "training model." This model is then used for prediction based on the test inputs provided to this trained prediction model (D'Amico *et al.,* 2019).

Few studies have explored various ML models and data mining applications for environmental assessment, particularly in the building sector. For instance, Azari *et al.,* (2016) and Sharif and Hammad (2019) used a combination of neural networks and LCA for optimal design selection of building envelopes and renovation methods, respectively. Meanwhile, another study proposes the use of ANN for a green building assessment system based on LCA (Xia and Liu 2013). Galimshina *et al.,* (2019) used a probabilistic LCA using advanced statistical methods to identify renovation strategies for buildings. Although studies have employed the use of ANN and regression models for energy assessments (D'Amico *et al.,* 2019; Li *et al.,* 2019), limited research is found using surrogate/substitute models for performing a complete LCA of buildings. However, such predictive modelling to perform LCA is proposed and demonstrated in other sectors such as agriculture machinery (Ma and Kim, 2015), chemical industry (Calvo-Serrano *et al.,* 2018), and product design (Park and Seo, 2003).

Nevertheless, considering the complexity of size, scale, and long-life span of buildings, dynamically performing a whole-building LCA is often very time-consuming and challenging (Francis and Thomas, 2022c). Hence, it is evident that the building sector needs to embrace modern computational and statistical tools to improve the decision-making process while proposing environmental conscious development. A robust surrogate model used for prediction purposes is efficient if it can ensure minimal loss of information while enabling full implementation of LCA (Eddy *et al.,* 2014). Therefore, this study demonstrates the potential of multiple regression and ANN tools in creating models to serve as substitutes for performing LCA of buildings, thereby reducing the time, cost, and complexities associated with it. Such models could then help in easing the decision-making process during project planning. The following section describes the methodological framework to develop such substitute models to perform LCA of buildings.

METHOD

Figure 1 shows the generic methodology proposed in the paper to enable the development of surrogate models to perform LCA for buildings. The first step is to collect inventory data on several building projects in terms of material consumption, transport of materials, equipment use and electricity consumption during the entire life cycle of the building. Once the life cycle inventory is compiled for various buildings, it should be analysed further for environmental impacts using suitable standard impact assessment methods (for the specific geographical boundary under consideration) or by using experimental and analytical conversion/characterisation factors (Guinee *et al.,* 2011). Once the LCA impacts are generated in the form of different indicator values (global warming potential, human health, ecosystem quality, acidification potential, etc.), the data is pre-processed to eliminate duplicate, missing, or inconsistent data. The input (inventory data) and output (LCA indicators) data are

then used to develop the ML model for prediction after data pre-processing. Any suitable ML algorithm can be used to develop the training model. However, the prediction accuracy among various methods can vary. So, the suitable algorithm can be chosen depending on the area of application. Once trained with a suitable ML algorithm, the model is then tested and validated for accuracy. Post validation, the model created can be used to obtain LCA indicator results for any building without performing a detailed whole building LCA.

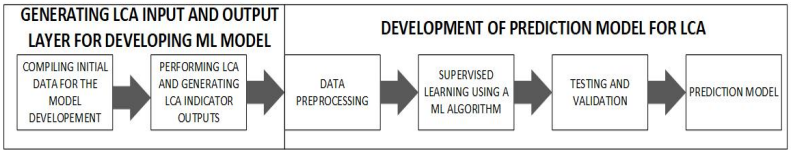

Figure 1: Research Method

The proposed methodology is demonstrated using a case dataset for Indian buildings. For this, an inventory data of 1000 buildings that contain information about five primary building materials used for building construction and their transportation to the site in ton-kilometres is collected. The five materials chosen are cement, sand, steel, bricks/AAC blocks and aggregate. The choice of five materials is made considering that these majorly affect life cycle energy and emissions according to studies in the Indian context (Reddy and Jagadish, 2003). This data is fed as inputs into the Simapro software (Version 9.1.0.8), which is software that assists in performing LCA. This data inventory is then assessed using a suitable impact assessment method. In this case, the ReCiPe end-point method (Hierachist Version) in association with the Ecoinvent database for these materials in Indian conditions is used. Post the inventory analysis using this method; results are obtained in the form of ReCiPe end-point indicators, namely Human health (Disability Adjusted Life Years (DALY)), Ecosystem Quality (Species. Year), and Resources (US Dollars) (Owsianiak *et al.,* 2014). The building inventory dataset of 1000 buildings and the three ReCiPe indicators from the impact assessment of each building becomes the training dataset for developing the ML model.

Accordingly, both multiple regression-based algorithms and an ANN-based ML model are experimented with to develop and compare prediction models based on the training dataset of 1000 buildings. A python code is used for the multiple regression model where the following relationship is derived from the model as shown in equation (1). Using these coefficients that are trained from the regression model, different LCA indicator values are predicted for a set of test data to check the robustness of the model generated.

$$Y=£_0+£_1X_1+£_2X_2+£_3X_3+£_4X_4+£_5X_5+£_6X_6+£_7X_7+£_8X_8+£_9X_9+£_{10}X_{10}+£_{11}X_{11}+€ \quad (1)$$

Where:

Y is the predicted results of ReCiPe end-point indicators.

X_1 to X_6 is the quantity of building materials in metric tonnes.

X_6 to X_{11} is the transportation involved for these five materials in ton-kilometres.

$£_0, £_1 \ldots £_{11}$ correspond to the regression coefficients

€ is the error variable.

Similarly, an ANN-based prediction model is also developed to compare the accuracy of using substitute or surrogate models for LCA. Figure 2 shows the framework of

ANN used in this study. For this, the 1000 building inventory and impacts dataset generated earlier is pre-processed or normalized if necessary and then divided into training data (80%) and testing data (20%). The ANN structure is defined by the input layer (input parameters same as that used for the regression model), output layer (values of LCA end-point indicators) and the hidden layers in-between. The input layer in this study includes 11 different variables (materials and transportation), as shown in Figure 2 and the output layer contains the three LCA end-point indicators under the ReCiPe method. The ANN structure is developed in MATLAB (Version 2020a) (as shown in Figure 2). The hidden layers are assigned weights that help initiate the activation function. This activation function maps the non-linear relationships between the input and output layers. This non-linear analysis is what distinguishes ANN from the multiple regression model. The transformation/training of data takes place in the hidden layers of neurons based on the nature of the data trained. Bayesian Regularisation is used in training the data. The iterative process of associating weights and bias to the model continues till the number of epochs specified is achieved or when the minimizing criteria such as the root mean squared errors (RMSE) is reached.

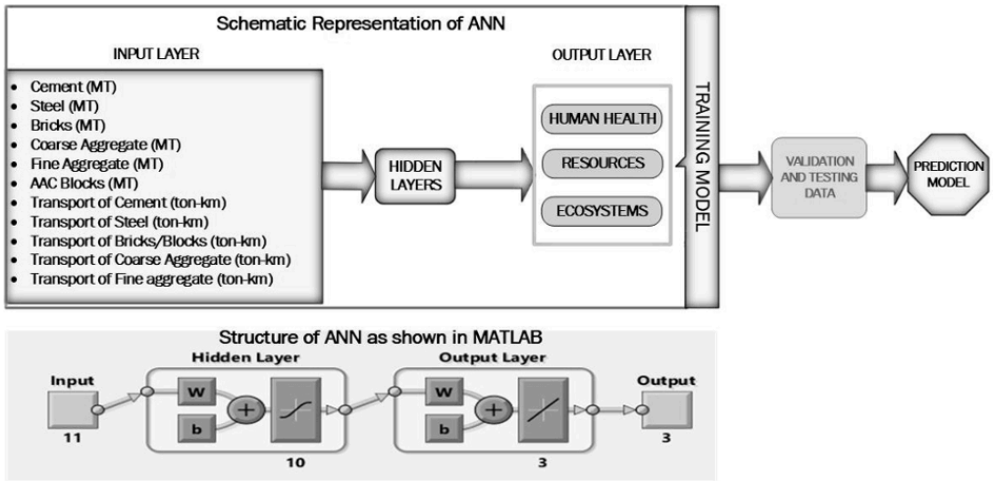

Figure 2: Framework of ANN interface

FINDINGS

Figure 3 shows the difference in percentage between the predicted versus actual results of ReCiPe end-point indicators for 10 test buildings for both regression and ANN. From these results, both these methods are closely representing the underlying structure of the LCA impact assessment method employed since the difference is within an acceptable range of less than 5%. Meanwhile, Figure 4 shows the results of the ANN training in MATLAB in terms of means squared error value rendering it a satisfactory prediction model. However, ANN is a more accurate prediction model when the data has a non-linear nature, as some existing studies conclude (Davis *et al.,* 2017; Ziyadi *et al.,* 2019).

Given the linear nature of the impact assessment method for LCA adopted in the case study, further extensive testing of the framework was done on the multiple regression model. The multiple regression models are tested with about 100 new building data samples. The mean absolute percentage error (MAPE) indicates the level of accuracy in terms of the difference between predicted and actual.

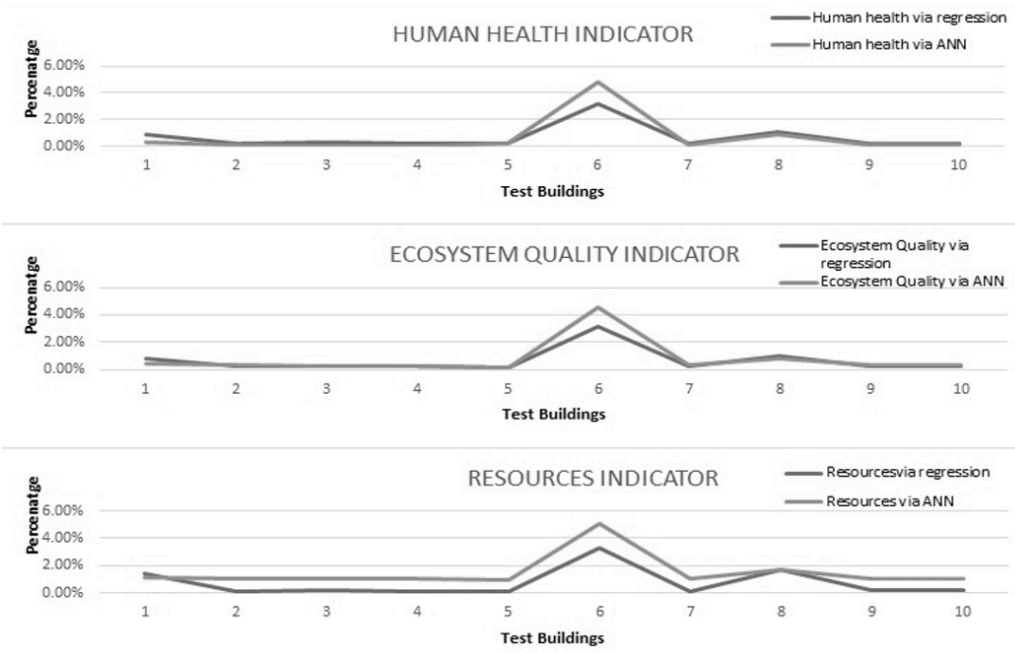

Figure 3: Regression vs ANN comparison of the percentage difference between predicted and actual

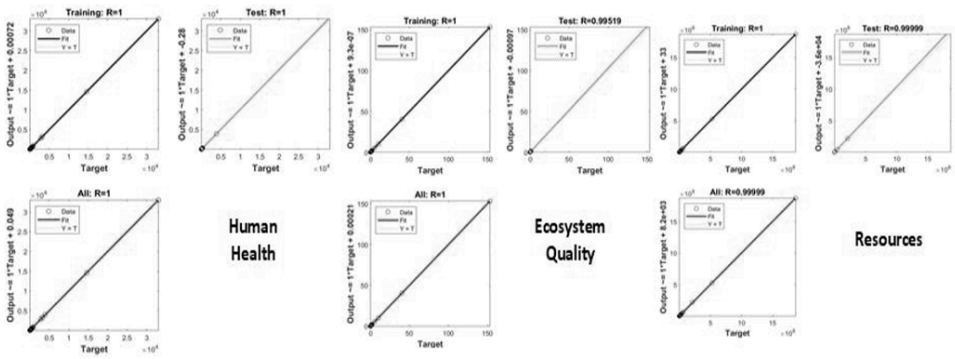

Figure 4: Results of ANN training

Therefore, for end-point indicators of human health, ecosystem quality, and resources, the MAPE for the samples tested is 4.46%, 4.52%, and 4.71%, respectively, as shown in Figure 5. It is a reasonably good model since the MAPE is below 5%. Hence, it can be used further in other decision-making frameworks for buildings which is demonstrated in the paragraphs further.

When building inputs vary (electricity consumption) constantly with time (dynamically), performing LCA becomes further tedious. Similarly, when LCA is adopted for decision-making in evaluating materials and their sourcing alternatives, it complicates the process of performing multiple scenarios with multiple datasets (Francis and Thomas, 2022b). For such dynamic applications, the use of substitute prediction models based on ML for performing LCA eases the process of decision-making even if their outputs are an approximate representation of actual cases. For instance, if there is a variation or replacement in the material content, such as cement or steel, remodelling and comparison can be quickly made using the surrogate/proxy LCA model.

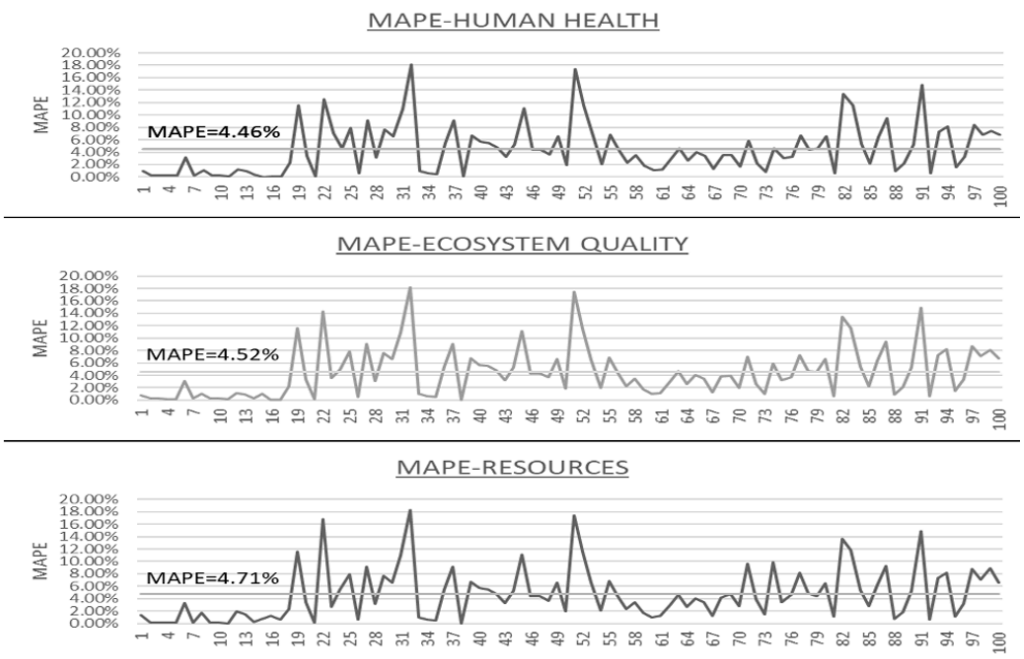

Figure 5: Mean Absolute Percentage Error for LCA indicators using regression

Figure 6 shows the average change in ReCiPe indicators with respect to the base case when material content is modified, as shown as percentage replacement from the base case. The graph shows the comparison between the actual LCA results of each scenario with the results generated from the substitute prediction LCA model as well. It shows that the trends are similar and data points are close for the 40 scenarios analysed.

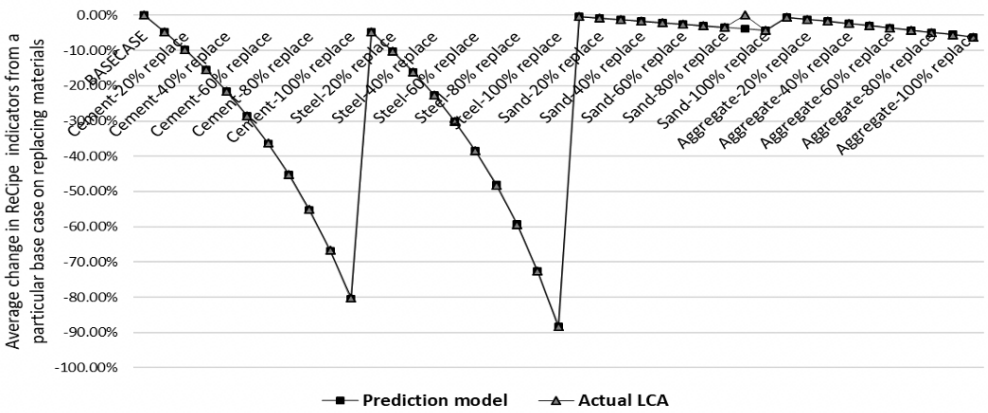

Figure 6: Average Change in LCA indicators with material variation

The advantage of using proxy LCA models is the ability to reduce complex computing when various scenario analyses are involved in enabling decision-making regarding environmental sustainability. For instance, in this specific building test dataset, the materials are replaced individually from 0 to 100 percent to evaluate the change in LCA indicators. By using the ML model, the results are obtained swiftly, and it can be observed that the cement and steel replacement significantly reduced the end-point indicators. These materials have the highest environmental impacts in terms of energy consumption among building materials, as reported by Bardhan (2011) and Jyosyula *et al.,* (2020). For this test data also, similar results are observed. Such analysis could enable the formulation of suitable policies such as local sourcing of materials to

reduce trip emissions or policies regarding the use of alternative materials with lower environmental impacts. Hence, the use of substitute prediction models for LCA is recommended to enhance the decision-making process while overcoming the challenges of data gaps, data rigor, time consumption and non-uniformity in assessment methods.

CONCLUSIONS

The study introduces the application of ML algorithms to perform LCA analysis of buildings which is conventionally a time-consuming, data-intensive, and rigorous process due to the scale and size of the data involved. Further, when LCA-based decision-making is based on several scenario analyses, conducting LCA for each scenario is complex. Here, a substitute/proxy prediction model would efficiently help predict the LCA-based indicators for the buildings. Therefore, this study proposes a methodological framework to develop ML-based surrogate models to perform LCA for buildings. Testing such models on a sample building dataset shows that the MAPE between the actual and predicted is less than 5%, rendering such models suitable for prediction. However, a limitation of such ML models is the need to constantly update training data as common background LCA databases are updated periodically. Further, it would be advisable to use primary data regarding environmental impacts and characterisation factors of materials from actual sources to account for geographical influence on the data inventory and, eventually, the training dataset. The future scope should focus on testing such ML frameworks by generating larger training datasets with more materials, electricity consumption, and other entities associated with building such that a whole building LCA can be entirely done using these substitute ML-based models, enabling faster scenario analysis and its improved use as a decision-making tool.

REFERENCES

Azari, R, Garshasbi, S, Amini, P, Rashed-Ali, H and Mohammadi, Y (2016) Multi-objective optimisation of building envelope design for life cycle environmental performance, *Energy and Buildings*, **126**, 524-534.

Bardhan, S (2011) Embodied energy analysis of multi storied residential buildings in urban India, *WIT Transactions on Ecology and the Environment*, **143**, 411-421.

Barros, N N and Ruschel, R C (2020) Machine Learning for Whole-Building Life Cycle Assessment: A Systematic Literature Review, *In*: E Toledo Santos and S Scheer (Eds.) *Proceedings of the 18th International Conference on Computing in Civil and Building Engineering ICCCBE 2020*, Lecture Notes in Civil Engineering, **98**, Cham: Springer.

Beloin-Saint-Pierre, D, Albers, A, Hélias, A, Tiruta-Barna, L, Fantke, P, Levasseur, A, Benetto, E, Benoist, A and Collet, P (2020) Addressing temporal considerations in life cycle assessment, *Science of the Total Environment*, **743**, 140700.

Cabeza, L F, Rincón, L, Vilariño, V, Pérez, G and Castell, A (2014) Life cycle assessment (LCA) and life cycle energy analysis (LCEA) of buildings and the building sector: A review, *Renewable and Sustainable Energy Reviews*, **29**, 394-416.

Calvo-Serrano, R, Gonzalez-Miquel, M, Papadokonstantakis, S and Guillén-Gosálbez, G (2018) Predicting the cradle-to-gate environmental impact of chemicals from molecular descriptors and thermodynamic properties via mixed-integer programming, *Computers and Chemical Engineering*, **108**, 179-193.

Chau, C K, Leung, T M and Ng, W Y (2015) A review on life cycle assessment, life cycle energy assessment and life cycle carbon emissions assessment on buildings, *Applied Energy*, **143**, 395-413.

D'Amico, A, Ciulla, G, Traverso, M, Lo Brano, V and Palumbo, E (2019) Artificial Neural Networks to assess energy and environmental performance of buildings: An Italian case study, *Journal of Cleaner Production*, **239**, 117993.

Davis, S E, Cremaschi, S and Eden, M R (2017) Efficient surrogate model development: optimum model form based on input function characteristics, *Computer-Aided Chemical Engineering*, 457-462.

Eddy, D, Krishnamurty, S, Grosse, I, Wileden, J and Lewis, K (2014, August) A robust surrogate modelling approach for material selection in sustainable design of products *In*: *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, American Society of Mechanical Engineers, **46285**, V01AT02A091.

Fouquet, M, Levasseur, A, Margni, M, Lebert, A, Lasvaux, S, Souyri, B, Buhé, C and Woloszyn, M (2015) Methodological challenges and developments in LCA of low energy buildings: Application to biogenic carbon and global warming assessment, *Building and Environment*, **90**, 51-59.

Francis, A and Thomas, A (2020) Exploring the relationship between lean construction and environmental sustainability: A review of existing literature to decipher broader dimensions, *Journal of Cleaner Production*, **252**, 119913.

Francis, A and Thomas, A (2022a) A framework for dynamic life cycle sustainability assessment and policy analysis of built environment through a system dynamics approach, *Sustainable Cities and Society*, 103521.

Francis, A and Thomas, A (2022b) System dynamics modelling coupled with multi-criteria decision-making (MCDM) for sustainability-related policy analysis and decision-making in the built environment, *Smart and Sustainable Built Environment, Emerald Publishing Limited,* [ahead of print].

Francis, A and Thomas, A (2022c) *A System Simulation Framework to Evaluate the Sustainability of Buildings In Infrastructure Development-Theory, Practice and Policy*, London: Routledge,129-141.

Galimshina, A, Hollberg, A, Moustapha, M, Sudret, B, Favre, D, Padey, P, Lasvaux, S and Habert, G (2019) Probabilistic LCA and LCC to identify robust and reliable renovation strategies, *IOP Conference Series: Earth and Environmental Science*, IOP Publishing, 12058.

Hellweg, S and Milà i Canals, L (2014) Emerging approaches, challenges and opportunities in life cycle assessment, *Science*, **344**(6188), 1109-1113.

Jyosyula, S K R, Surana, S and Raju, S (2020) Role of lightweight materials of construction on carbon dioxide emission of a reinforced concrete building, *Materials Today*, 27, 984-990.

Levasseur, A, Lesage, P, Margni, M, Deschěnes, L and Samson, R (2010) Considering time in LCA: Dynamic LCA and its application to global warming impact assessments, *Environmental Science and Technology*, **44**(8), 3169-3174.

Li, Z, Dai, J, Chen, H and Lin, B (2019) An ANN-based fast building energy consumption prediction method for complex architectural form at the early design stage, *Building Simulation*, 665-681.

Ma, J and Kim, H M (2015) Predictive usage mining for life cycle assessment, *Transportation Research Part D: Transport and Environment*, **38**, 125-143.

Morales, M, Moraga, G, Kirchheim, A P and Passuello, A (2019) Regionalized inventory data in LCA of public housing: A comparison between two conventional typologies in southern Brazil, *Journal of Cleaner Production*, **238**, 117869.

Owsianiak, M, Laurent, A, Bjørn, A and Hauschild, M Z (2014) IMPACT 2002+, ReCiPe 2008 and ILCD's recommended practice for characterisation modelling in life cycle impact assessment: A case study-based comparison, T*he International Journal of Life Cycle Assessment*, **19**(5), 1007-1021.

Park, J H and Seo, K.-K (2003) Approximate life cycle assessment of product concepts using multiple regression analysis and artificial neural networks, *KSME International Journal*, **17**(12), 1969-1976.

Rebitzer, G, Ekvall, T, Frischknecht, R, Hunkeler, D, Norris, G, Rydberg, T, Schmidt, W.-P, Suh, S, Weidema, B P and Pennington, D W (2004) Life cycle assessment: Part 1: Framework, goal and scope definition, inventory analysis and applications, *Environment International*, **30**(5), 701-720.

Reddy, B V V and Jagadish, K S (2003) Embodied energy of common and alternative building materials and technologies, *Energy And Buildings*, **35**(2), 129-137.

Sharif, S A and Hammad, A (2019) Developing surrogate ANN for selecting near-optimal building energy renovation methods considering energy consumption, LCC and LCA, *Journal of Building Engineering*, **25**(March), 100790.

Sharma, A, Saxena, A, Sethi, M and Shree, V (2011) Life cycle assessment of buildings: a review, *Renewable and Sustainable Energy Reviews*, **15**(1), 871-875.

Sousa, I, Wallace, D, Borland, N and Deniz, J (1999) A learning surrogate LCA model for integrated product design 1 Introduction, *Proceedings of the 6th International Seminar on Life Cycle Engineering*, 209-219.

Version 9.1.0.8 Simapro 9.1.0.8 [Computer Software] PRe consultants (2011), The Netherlands.

Xia, L M and Liu, J J (2013) *Research On Green Building Assessment System Based On Bp Neural Network And Life Cycle Assessment (LCA) Applied Mechanics And Materials*, Trans Tech Publications Ltd, 508-514.

Ziyadi, M and Al-Qadi, I L (2019) Model uncertainty analysis using data analytics for life-cycle assessment (LCA) applications, *The International Journal of Life Cycle Assessment*, **24**(5), 945-959.